Future Work on Graph Algorithms

Bruce Hendrickson
Edmond Chow
Tina Eliassi-Rad
William McLendon
Keith Henderson
Andy Yoo

Work supported by the DHS ASC Program

Multifaceted Agenda

- Connection Subgraphs
  - Move beyond short paths
- Parallel Partitioning
  - Enable parallel graph algorithms
- Subgraph Isomorphism
  - Find instance of *interesting* graph in dataset
- Alternative Architecture Exploration
  - Blue Gene / Light
  - Massively multithreaded machines

Beyond Short Paths

- Short paths are a means to an end:
  - Find interesting relations, connections & communities
- Some paths more interesting than others
  - E.g. avoid high-degree intermediate nodes, and less interesting edge types
- *Connection Subgraphs* are small graphs that best describe relationships between two entities
  - Uses circuit metaphor (Faloutsos, et al. 2004)

Connection Subgraphs

- Implement parallel algorithm for connection subgraphs
  - Start with subgraph of all short paths
- Extend existing models to handle network dynamics
  - E.g. directed edges or temporal considerations

Parallel & Dynamic Partitioning

- Current tools assume graph is partitioned serially.
  - Not possible for huge or changing graphs
- Zoltan toolkit includes suite of parallel partitioners
- Zoltan being integrated with CompNets
  - One-dimensional partitioning implemented
  - Two-dimensional methods in progress

Motif Finding: Example

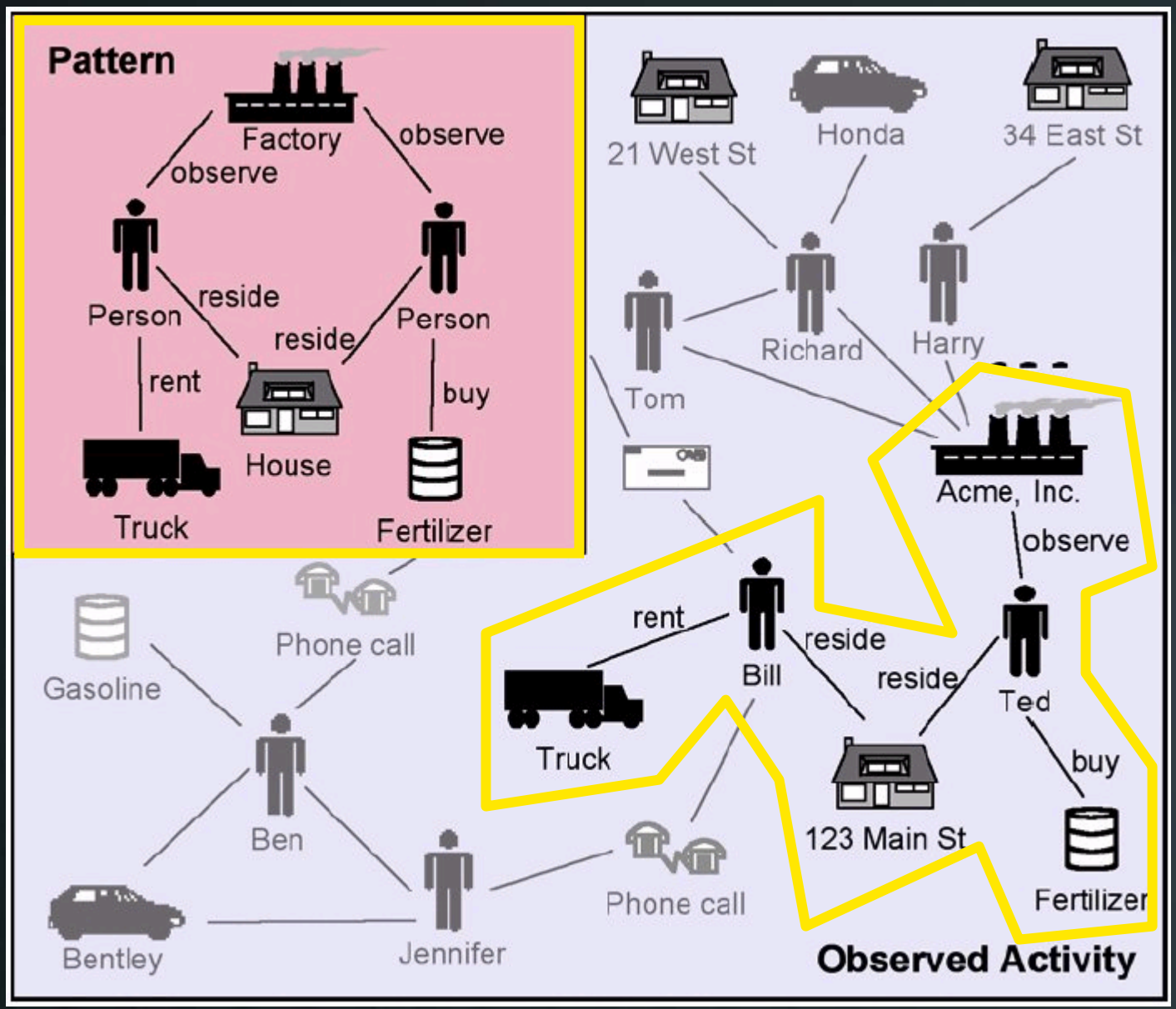

Image Source:
T. Coffman, S. Greenblatt, S. Marcus, *Graph-based technologies for intelligence analysis*, CACM, 47 (3, March 2004): pp 45-47

Finding Subgraphs

- Analyst might have a model of an *interesting* set of relationships
  - How do you search the graph to find instances?
  - Parallel algorithms for *subgraph isomorphism*

- Problem NP-Hard, but semantic structure helps
- We will develop parallel algorithms for very large instances

Alternative Architectures

- Atypical computers have potential appeal
- Blue Gene / Light
  - Lots of inexpensive computational power
- Cray MTA & Eldorado
  - Multithreading masks latency, which dominates cost of graph algorithms

- We are experimenting with these alternatives